

A Method of Combination of Language Understanding with Touch-Based Communication Robots

Takuki Ogawa, Hiroaki Bando, Kazuhiro Morita, Masao Fuketa, Jun-Ichi Aoe

University of Tokushima, Tokushima City, Japan

Email: ogawa-takuki@iss.tokushima-u.ac.jp, h-bando@aoelab.is.tokushima-u.ac.jp, kam@is.tokushima-u.ac.jp,
fuketa@is.tokushima-u.ac.jp, aoe@is.tokushima-u.ac.jp

Received July 23, 2012; revised August 25, 2012; accepted September 7, 2012

ABSTRACT

Studies of robots which aim to entertain and to be conversational partners of the live-alone become very important. The robots are classified into DBC (Dialogue-Based Communication) robots and TBC (Touch-Based Communication) robots. DBC robots have an effect to be conversational partners. A typical application of TBC robots is Paro (a baby harp seal robot) which has an effect to entertain humans. The combination of DBC and TBC will be able to achieve both a conversational ability and an entertaining effect, but there is no study of combination of DBC and TBC. This paper proposes a response algorithm that can combine conversational information and touch information from humans. Criteria for estimation are defined as follows: FFV (Familiarity Factor Value), EFV (Enjoyment Factor Value), CR (Concentration Rate), ER (Expression Rate), and RR (Recognition Rate). FFV and EFV are total ratings for questionnaire related to familiarity and enjoyment factors, respectively. CR measures attention for humans. ER is the interest of communication with robots by representing Level 2 (laugh with opening one's mouth), Level 1 (smile), and Level 0 (expressionless). RR is recognition ability for voices and touch actions. From the experiment for impressions of robot responses with 11 subjects, it turns out that the proposed method with combination of DBC and TBC is improved by 20.7 points in FFV, and by 12.6 points in EFV compared to only TBC. From the robot communication experiment, it turns out that the proposed method is improved by 8 points in the ER, by 5.3 points in the ER with Level 2, and by 24.5 points in the RR compared to only DBC.

Keywords: Natural Language Interface; Robot Communication; Touch-Based Dialogue; Empathetic Responses

1. Introduction

Studies of robots which aim to ease stress and to be conversational partners of the live-alone become very important. Application fields of these studies include the nursing care, rehabilitation, conversational partners of the live-alone and so on. In the robots, there are DBC (Dialogue-Based Communication) robots and TBC (Touch-Based Communication) robots. DBC robots have an effect to be conversational partners [1-3]. TBC robots such as Paro have an effect to entertain humans [4-8]. The combination of DBC and TBC will be able to achieve both a conversational ability and an entertaining, but there is no study of combination of DBC and TBC.

Although there are traditional studies of touch sensors for communication robots [9,10] and DBC robots with touch sensors [3], these studies don't mention combination of DBC and TBC.

In order to solve the problem, this paper proposes a method of combination of DBC and TBC and an empathetic communication scheme for compact stuffed robots controlled by Internet servers [11,12]. In order to esti-

mate empathetic communication schemes, the new empathy measurements for generalized conversations are defined. In the proposed method, the main empathy communication modules are located on Internet Server systems (IS) and the robot communication is controlled by this server. The IS enables us to reduce the size of robots and to remove many software modules from them. The robots consist of a microphone, a speaker and six pressure sensors perceiving by human touches.

Section 2 introduces the concept of robot communication systems. Section 3 describes outlines of the conversation systems. Section 4 proposes the classification of empathy for generalized conversations. Section 5 evaluates the proposed method. Section 6 concludes the proposed method and discusses future works.

2. The Concept of Robot Communication Systems

2.1. Traditional Studies

In the DBC systems, there are studies of task-oriented

dialogue technologies. R. Freedman [13] developed APE (the Atlas Planning Engine) which is the reactive dialogue planner for intelligent tutoring systems. B. S. Lin *et al.* [14] developed a method for multi-domain DBC systems. The method has different spoken dialogue agents (SDA's) for different domains and a user interface agent to access correct SDA using a domain switching protocol. But these task-oriented DBC systems have no function to assist conversations or to be conversational partners. The assisting feature is to inform conversational topics helping conversations between humans. The partners are to recognize conversational sentences from humans and reply appropriate responses for less-task oriented conversations such as chats between a human and a computer.

Next, consider studies of assist conversation systems. N. Alm *et al.* [15] indicated that people with dementia take a more active part in conversation to show their memories as conversation topics, and developed CIRCA (the Computer Interactive Reminiscence and Conversation Aid) that has a touch screen and presents user's reminiscence materials. A. J. Astell *et al.* [16] verified that CIRCA promotes and maintains conversations between people with dementia and caregivers. These studies are required to the function to be conversational partners.

As the DBC studies of conversational partners, there are DBC agents by Nakano [1]. First, the DBC agents ask questions. Second, the DBC agents nod and give positive answers for humans' speeches to the question. The agents use verbal acknowledgement based on the acoustic information such as the pitch and the intensity. Meguro *et al.* [17] analyzed dialogue flows of listening-oriented dialogues and casual conversations, and showed the difference of them for listening agents using DBC. However, the purpose of these systems is not interlacing DBC with TBC.

In the TBC systems, Shibata [18] developed Paro whose appearance resembles a baby harp seal with a soft touch fur. The robot can have TBC by 6 ubiquitous surface tactile sensors and recognize human's voices by 3 microphones. Studies [6,7,19] of robot therapy using Paro indicated that humans improve their feelings by communications. In order to improve human's feelings, Kanoh [8] developed Babyloid robot behaving like a baby that the abdomen and hip are made from silicon with a soft touch. It can recognize TBC by 5 touch sensors and 2 optical sensors, and can recognize voices from humans by a microphone. These TBC robots have no ability to understand linguistic information, although they can recognize humans' voices.

Mitsunaga [3] developed Robovie-IV that is able to have rule-based DBC and to recognize touch actions from humans by 56 tactile sensors. In the study, it is necessary to decide responses by combination of linguis-

tic information and touch information.

2.2. A DBC Robot with TBC

Figure 1 shows the construction of the proposed system. In **Figure 1**, the system consists of a robot and a dialogue management server. The robot has 3 modules as follows; voice recognition, a sensor management and a sound play module. The voice recognition module converts human voices from microphone to voice strings. The sensor management module gets raw signals from each touch sensor, and detects touch positions by threshold processing. The sound play module produces robot voices from voice data.

The dialogue management server has 3 modules; an intention estimate, a decide response and a voice synthesis module. The intention estimate module judges 81 kinds of speech intentions from voice strings. The decide response module generates response strings using dialogue knowledge from a speech intention and a touch position. The voice synthesis module converts response strings to voice data.

Figure 2 shows a DBC robot with TBC. In **Figure 2**, the robot uses the stuffed animal because it is soft to touch. Moreover, the robot has six touch sensors in its body; one in the head, four in the left and right limbs and one in the abdomen. One microphone and one speaker are equipped behind the robot. It is a little difficult to recognize voices by a microphone than by a headset, but it enables us to have natural dialogues between human and the robot.

Tables 1 and **2** show examples of only DBC and the combination of DBC and TBC, respectively. These figures indicate humans' answers for the robot request "I have a headache. Please stroke my head" and robot responds for the answers. In **Table 1** and **2**, attributes for humans' answers are defined by <AFFIRMATIVE>, <NOTHING>, <NEGATIVE> and <OTHERS>. <AFFIRMATIVE> means answer yes such as "O.K. I will do it". In this case, the robot replies thankful responses such as "Thanks, how nice of you. I can forget a headache." in **Figure 3**. <NOTHING> means no answer. In this case, the robot asks whether humans hear the requests or not. For example, "Oh, you might not hear." in **Figure 3**. <NEGATIVE> means gainsaying the requests such as "I don't want to touch". In this case, the robot expresses sad intentions such as "Don't you want to touch me? I'm sad." in **Figure 3**. <OTHERS> is the answer, which has no relation to the request, such as "I am tired." In this case, the robot asks if humans have the intentions to accept the requests. The example is "You don't feel like touching me, do you?" in **Table 1**.

Attributes for touch actions are defined by <TOUCHED>, <MIS-TOUCHED> and <NON-TOUCHED>.

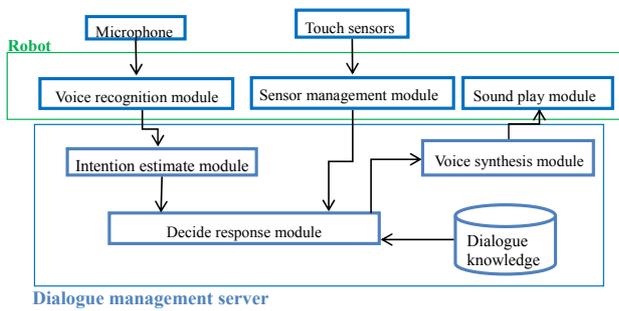


Figure 1. The construction of the proposed system.

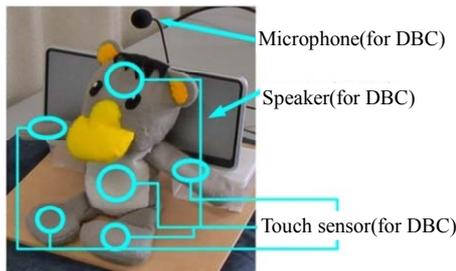


Figure 2. A DBC robot with TBC.

<TOUCHED> means humans made physical contacts with the proper position requested by robots. <MIS-TOUCHED> means humans made physical contacts with outside of the requested position. <NON-TOUCHED> means humans didn't make physical contacts with the robot.

Although Table 1 defines only four types of robot responses to humans' answer attributes, but Table 2 can define twelve types of robot responses to attributes of humans'

Table 1. Examples for attributes of humans' answers and robot responses.

Attributes for humans' answers	Robot responses
<AFFIRMATIVE> ("O.K. I will do it")	"Thanks, how nice of you. I can forget a headache"
<NOTHING> (No answer)	"Oh, you might not hear."
<NEGATIVE> ("I don't want to touch.")	"Don't you want to touch me? I'm sad."
<OTHERS> ("I am tired.")	"You don't feel like touching me, do you?"

Table 2. Examples of the combination of DBC and TBC robot responses.

Attributes for touch actions Attributes for humans' answers	<TOUCHED> (Touched robot's head)	<MIS-TOUCHED> (Touched robot's body outside head. For example, left hand)	<NON-TOUCHED>
<AFFIRMATIVE> ("O.K. I will do it")	"Thanks, how nice of you. I can forget a headache."	"Thank you for trying to touch me."	"Thanks, how nice of you. I can forget a headache"
<NOTHING> (No answer)	"I think I feel better. Thank you."	"That is not the place I want, but I'm happy."	"Oh, you might not hear."
<NEGATIVE> ("I don't want to touch.")	"I'm happy because you touch me, though."	"I knew you are sweet. I'm happy."	"Don't you want to touch me? I'm sad."
<OTHERS> ("I am tired.")	"My headache has gone because you touched me."	"You touch my left hand, but you make me happy by your action."	"You don't feel like touching me, do you?"

answers and touch actions. These robot responses are explained as follows:

Case 1) <TOUCHED> and <AFFIRMATIVE>

In this case, humans express strong intentions to accept the requests by attributes of the humans' answers and the touch actions. Therefore, the robot can reply "very thankful" answers such as "Thanks, how nice of you. I can forget a headache." in Table 2.

Case 2) <TOUCHED> and <NOTHING>

The proper result can be obtained from the touch action attribute, even if the humans' answer attribute is <NOTHING>. Therefore, the robot can reply "thankful" answers such as "I think I feel better. Thank you." in Table 2.

Case 3) <TOUCHED> and <NEGATIVE>

Although humans say the negative answer, the robot

can reply "happy" answers such as "I'm happy because you touch me, though" in Table 2 by the touch action attribute.

Case 4) <TOUCHED> and <OTHERS>

Humans express an intention to accept the request by the touch action attribute even if the humans' answer attribute has no relation to the request. Therefore, the robot can reply "thankful" answers such as "My headache has gone because you touched me." in Table 2.

Case 5) <MIS-TOUCHED> and <AFFIRMATIVE>

Humans touch parts except the requested position by the robot, but humans agree with the robot request.

The robot can reply "thankful" responses such as "Thank you for trying to touch me." in Figure 4 by reflecting the <MIS-TOUCHED> attribute.

Case 6) <MIS-TOUCHED> and <NOTHING>

The humans' answer attribute has no information but there is the fact that human tried to touch parts of the robot. Therefore, the robot can reply "happy" responses such as "That is not the place I want, but I'm happy." in **Figure 4** by reflecting the <MIS-TOUCHED> attribute.

Case 7) <MIS-TOUCHED> and <NEGATIVE>

The humans' answer attribute has opposite intentions for the requests, however there is the fact that human tried to touch parts of the robot as Case 6. Therefore, the robot can reply "a little happy" responses such as "I knew you are sweet. I'm happy." in **Figure 4** by reflecting the <MIS-TOUCHED> attribute.

Case 8) <MIS-TOUCHED> and <OTHERS>

Although the humans' answer has no relation to the request, humans tried to touch parts of the robot as Case 6. Therefore, the robot can express "happy" responses such as "You touch my left hand, but you make me happy by your action" in **Table 2**.

Case 9) <NON-TOUCHED> and <AFFIRMATIVE>

There is possibility that the robot failed to recognize touch actions from humans because humans accept requests. Therefore, the robot can reply "thankful" responses such as "Thanks, how nice of you. I can forget a headache." in **Table 2**.

Case 10) <NON-TOUCHED> and <NOTHING>

Since humans don't express any information and humans have no touch action, the robot asks whether humans hear the request or not. For example, "Oh, you might not hear." in **Table 2**.

Case 11) <NON-TOUCHED> and <NEGATIVE>

Humans oppose the request and humans have no touch action, thus the robot expresses "sad" intentions such as "Don't you want to touch me? I'm sad." in **Table 2**.

Case 12) <NON-TOUCHED> and <OTHERS>

The humans' answer has no relation to the request and humans have no touch action, therefore, the robot asks if humans have the intentions to accept the requests. The example is "You don't feel like touching me, do you?" as shown in **Figure 4**.

This study classifies human's answers with the combination of DBC and TBC, and proposes a response algorithm for each case in Section 3.

3. Outlines of the Conversation Systems

In the proposed system, a passive conversation scheme is introduced to keep coherent and empathy conversation because free conversations are not only difficult but also unpractical. Therefore, a question intention called a conditional question from the system restricts human's responses.

Definition 3.1. A conditional question is defined by question intention c_{qi} and condition $CONDITION(c_{qi})$ for c_{qi} where c_{qi} is represented by hierarchy of concepts, and

$CONDITION(c_{qi})$ is a set of concepts to be expected on humans' answers.

Definition 3.2. For question intention c_{qi} , $SPEECH(c_{qi})$ is defined as the function that generates speech sentences from c_{qi} .

For example, question intention $c_{qi} = \langle QUESTION \rangle / \langle TEMPERATURE \rangle / \langle COLD \rangle$, where $\langle \rangle$ denotes the concept and / means the path of concepts or hierarchy. c_{qi} represents a question about temperature; whether "human is cold" or not. $SPEECH(c_{qi})$ generates questions such as "Is it cold now?". For c_{qi} , $CONDITION(c_{qi})$ is defined by a set of concepts <AFFIRMATIVE>, <NEGATIVE>, <NOTHING>, and <OTHERS>, where <AFFIRMATIVE> means the concept of affirmative answers such as "yes", <NEGATIVE> means the concept of negative answers such as "no", and <NOTHING> means "no answer". These three concepts are related to restricted and expected answers. However, <OTHERS> means the unrestricted concept of answers except the above three concepts and the answers are understood by the generalized intension engine.

Consider a touch-based conditional question as follows:

$c_{qi} = \langle REQUEST \rangle / \langle TOUCH \rangle / \langle HEAD \rangle$

$CONDITION(c_{qi}) = \{ \langle TOUCHED \rangle, \langle MIS-TOUCHED \rangle, \langle NON-TOUCHED \rangle, \langle AFFIRMATIVE \rangle, \langle NEGATIVE \rangle, \langle NOTHING \rangle \text{ and } \langle OTHERS \rangle \}$

One of questions by $SPEECH(c_{qi})$ is "I have a headache. Please stroke my head."

<TOUCHED> means that the head has been touched and the human action is complete. <MIS-TOUCHED> means that other positions instead of the head have been touched and the human action is not complete. <NON-TOUCHED> means no positions have been touched.

For requests from the robot, the human returns utterance u and action a . Then, $MEANING(u)$ and $MEANING(a)$ represent intention meanings for u and a , respectively. Response sentences are determined by matching between a concept of conditional questions and a concept of intension meanings. The matching is defined by the function, MATCH

The formal conversation algorithm is proposed as follows:

[Conversation Algorithm]

Input: Robot intention c_{qi} . $CONDITION(c_{qi})$ is a set of concepts for c_{qi} .

Output: Response sentences r for c_{qi} .

Method:

(Step 1): {Robot speech generation.}

Speak $SPEECH(c_{qi})$ to the human.

(Step 2): {Decision about whether the speech intension is touch-base request, or not.}

If c_{qi} is touch-based request go to **(Step 4)**, otherwise go to the next **(Step 3)**.

(Step 3): {Processing of the DBC}

Suppose that u is the utterance from the human.

Compute $\text{MEANING}(u)$ and set the result $[c_{u1}, c_{u2}]$ to

C_u :

For all pairs $C = [c, c_{qi}]$ for c in $\text{CONDITION}(c_{qi})$,

determine the candidate C_{u1} by $\text{MATCH}(C_u, C)$:

Set $\text{REPLY}(C_{u1})$ to r and terminate this algorithm.

(Step 4): {Determination of intension for the TBC}

Suppose that u is the utterance and a is the action from the human.

Compute $\text{MEANING}(u)$ and set the result $[c_{u1}, c_{u2}]$ to

C_u :

Compute $\text{MEANING}(a)$ and set the result $[c_{a1}, c_{a2}]$ to

C_a :

(Step 5): { Response generation for the touch-based conversation}

For all pairs $C = [c, c_{qi}]$ for c in $\text{CONDITION}(c_{qi})$,

determine the candidate C_{u1} by $\text{MATCH}(C_u, C)$ and determine the candidate C_{a1} by $\text{MATCH}(C_a, C)$:

By using $\text{PRIORITY}(C_{u1}, C_{a1})$, set $\text{REPLY}(C_{a1})$ and $\text{REPLY}(C_{u1})$ to r :

Terminate this algorithm:

End of Algorithm

Figures 3 and 4 show the processing flow of the conversational algorithm for DBC and TBC, respectively.

Example 3.1. Consider the above example $c_{qi} = \langle \text{QUESTION} \rangle / \langle \text{TEMPERATURE} \rangle / \langle \text{COLD} \rangle$

$\text{CONDITION}(c_{qi}) = \{ \langle \text{AFFIRMATIVE} \rangle, \langle \text{NEGATIVE} \rangle, \langle \text{NOTHING} \rangle, \langle \text{OTHERS} \rangle \}$

(Step 1)

Speech sentence is generated by $\text{SPEECH}(c_{qi}) = \text{“Is it cold now?”}$

(Step 2)

c_{qi} is not touch-based request then go to (Step3)

(Step 3)

Suppose that human utterance for the speech sentence is $u = \text{“I feel cold”}$, and that $\text{MEANING}(u)$ is $C_u = [\langle \text{FEEL} \rangle, \langle \text{COLD} \rangle]$. For C_u and for all pairs $C = [c, c_{qi}]$ for c in $\text{CONDITION}(c_{qi})$, the function $\text{MATCH}(C_u, C)$ determines $C_{u1} = [\langle \text{AFFIRMATIVE} \rangle, c_{qi}]$ because $\langle \text{FEEL} \rangle$ is not negative and the concept of c_{qi} includes $\langle \text{COLD} \rangle$. Then, $\text{“It is better to turn on heating.”}$ is obtained as an example of $\text{REPLY}(C_{u1})$.

Table 3 shows examples of response sentences for $u = \text{“Is it cold?”}$ In **Table 1**, $u = \text{NULL}$ means no answer from humans and NULL is also utilized by denoting no concepts. $U = \text{“yes”}$ in the first case is the typical affirmative answer and $\text{MEANING}(u)$

becomes $[\langle \text{AFFIRMATIVE} \rangle, \text{NULL}]$.

For $u = \text{NULL}$ of the second case, although there is no answer, the proposed method produces the empathy response as $\text{REPLY}(C_{u1}) = \text{“Oh, you might not hear”}$. $u = \text{“I feel hot”}$ in the third case is the negative answer be

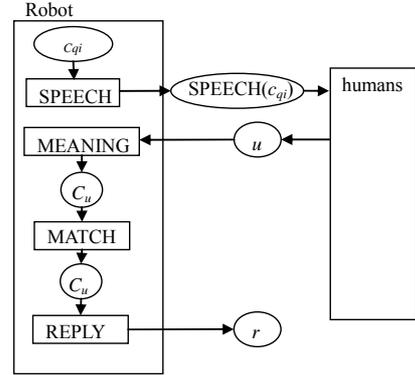


Figure 3. The processing flow of the conversational algorithm for DBC.

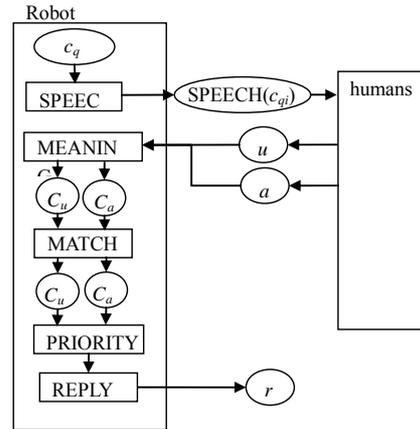


Figure 4. The processing flow of the conversational algorithm for TBC.

Table 3. Examples of response sentences for $u = \text{“Is it cold?”}$

u	$\text{REPLY}(C_{u1})$
$u = \text{“yes”, “sure”, ...}$	$u = \text{“yes”}$ $C_u = \text{MEANING}(u) = [\langle \text{AFFIRMATIVE} \rangle, \text{NULL}]$ $C_{u1} = [\langle \text{AFFIRMATIVE} \rangle, c_{qi}]$ $\text{REPLY}(C_{u1}) = \text{“Thanks, how nice of you.”}$
$u = \text{NULL}$ (no answer from human)	$u = \text{NULL}$ $C_u = \text{MEANING}(u) = [\langle \text{NOTHING} \rangle, \text{NULL}]$ $C_{u1} = [\langle \text{NOTHING} \rangle, c_{qi}]$ $\text{REPLY}(C_{u1}) = \text{“Oh, you might not hear.”}$
$u = \text{“I feel hot”, “no”, “it is hot”, “I don’t feel cold”}$	$u = \text{“I feel hot”}$ $C_u = \text{MEANING}(u) = [\langle \text{FEEL} \rangle, \langle \text{HOT} \rangle]$ $C_{u1} = [\langle \text{NEGATIVE} \rangle, c_{qi}]$ $\text{REPLY}(C_{u1}) = \text{“I think you are exciting now.”}$
$u = \text{“I’m tired now”, “Hungry”, “I have no money”, ...}$	$u = \text{“I’m tired now”}$ $C_u = \text{MEANING}(u) = [\langle \text{TIRED} \rangle, \text{NULL}]$ $C_{u1} = [\langle \text{OTHERS} \rangle, c_{qi}]$ $\text{REPLY}(C_{u1}) = \text{“Sleeping and eating is very important”}$

cause it includes the contrast word “hot” and “cold”. Then, MEANING(u) becomes [\langle FEEL \rangle , \langle HOT \rangle] and the empathy response REPLY(C_{ul}) = “I think you are exciting now” is produced. The fourth example u = “I have no money” has no relation to the question. The proposed method understands this sentence by the generalized intension recognition (GIR) modules which are independent of conditional questions. Therefore, the empathy response REPLY(C_{ul}) = “Sleeping and eating is very important” is produced.

Table 4 shows examples of response sentences such as “I have a headache. Please stroke my head”.

Table 4 has very important features in the empathy conversation because **Table 4** indicates the priority to responses for touch-based conversation.

More details are explained in Example 3.2.

Example 3.2. Consider the touch-based conversation by example c_{qi} = \langle REQUEST \rangle / \langle TOUCH \rangle / \langle HEAD \rangle CONDITION(c_{qi}) = { \langle AFFIRMATIVE \rangle , \langle NEGATIVE \rangle , \langle NOTHING \rangle , \langle OTHERS \rangle , \langle TOUCHED \rangle , \langle NON-TOUCHED \rangle , \langle MIS-TOUCHED \rangle }

(Step 1)

The speech sentence is generated by SPEECH(c_{qi}) = “I have a headache. Please stroke my head.”

(Step 2)

c_{qi} is a touch-based request, then go to (Step 4)

In order to explain Steps 4 and 5, Table 4 shows examples of c_{qi} = \langle REQUEST \rangle / \langle TOUCH \rangle / \langle HEAD \rangle . In general, touch results such as \langle TOUCHED \rangle , \langle MIS-TOUCHED \rangle and \langle NON-TOUCHED \rangle have priority over dialogue results such as \langle AFFIRMATIVE \rangle , \langle NEGATIVE \rangle , \langle NOTHING \rangle and \langle OTHERS \rangle . On the

other hand, the dialogue result \langle AFFIRMATIVE \rangle has priority of only the touch result \langle NON-TOUCHED \rangle . The proposed method defines the function PRIORITY(C_{ul} , C_{al}).

From the results of **Table 4**, it turns out that the proposed method is keeping the friendliness between the robot and the human.

Case 1: C_{al} = [\langle TOUCHED \rangle , c_{qi}]

In this case, the proper results are obtained for the touch requests. Therefore, REPLY(C_{al}) should response thankful and affective expressions. REPLY (C_{ul}) can be neglected because the TBC prioritizes the DBC.

Case 2: [\langle MIS-TOUCHED \rangle , c_{qi}]

\langle MIS-TOUCHED \rangle means the fact that a human tries to touch the position of the robot according to the touch request. Therefore, REPLY(C_{al}) also responses thankful and affective expressions with supplemental expressions about mis-touch information. By the same manner of \langle TOUCHED \rangle , REPLY (C_{ul}) can also be neglected.

Case 3: [\langle NON-TOUCHED \rangle , c_{qi}]

\langle NON-TOUCHED \rangle means that the touch devices have no information and it is impossible to understand the proper human actions. For example, we can consider humans have touched the position with no devices of the robot. Therefore, the responses should be used for dialogue information in this case.

4. Experimental Observations

In order to evaluate the proposed algorithm, two experiments are carried out; the one is the experiment for impressions of robot responses, and another is the robot

Table 4. Examples of response sentences for “I have a headache. Please stroke my head”.

PRIORITY(C_{ul} , C_{al})	C_{al} =[\langle TOUCHED \rangle , c_{qi}]	C_{al} =[\langle MIS-TOUCHED \rangle , c_{qi}] “Left-hand has been touched”	C_{al} =[\langle NON-TOUCHED \rangle , c_{qi}]
C_{ul} = [\langle AFFIRMATIVE \rangle , c_{qi}]	REPLY(C_{al}) = “Thanks, how nice of you. I can forget a headache” REPLY (C_{ul}) = “Thank you for trying to touch me.”	REPLY(C_{al}) = “It is my left hand, but I’m happy.” REPLY(C_{ul}) = “Thank you for trying to touch me.”	REPLY(C_{al}) = NULL REPLY(C_{ul}) = “Thanks, how nice of you. I can forget a headache.”
C_{ul} = [\langle NOTHING \rangle , c_{qi}]	REPLY(C_{al}) = “I think I feel better. Thank you.” REPLY (C_{ul}) = “Oh, you might not hear.”	REPLY(C_{al}) = “That is not the place I want, but I’m happy.” REPLY(C_{ul}) = “Oh, you might not hear.”	REPLY(C_{al}) = NULL REPLY(C_{ul}) = “Oh, you might not hear.”
C_{ul} = [\langle NEGATIVE \rangle , c_{qi}] “no”, “I don’t want to touch”	REPLY(C_{al}) = “I’m happy because you touch me, though” REPLY(C_{ul}) = “Don’t you want to touch me? I’m sad.”	REPLY(C_{al}) = “I knew you are sweet. I’m happy.” REPLY(C_{ul}) = “Don’t you want to touch me? I’m sad.”	REPLY(C_{al}) = NULL REPLY(C_{ul}) = “Don’t you want to touch me? I’m sad.”
C_{ul} = [\langle OTHERS \rangle , c_{qi}]	REPLY(C_{al}) = “My headache has gone because you touched me.” REPLY(C_{ul}) = “You don’t feel like touching me, do you?”	REPLY(C_{al}) = “You touch my left hand, but you make me happy by your action.” REPLY(C_{ul}) = “You don’t feel like touching me, do you?”	REPLY(C_{al}) = NULL REPLY(C_{ul}) = “You don’t feel like touching me, do you?”

communication experiment.

In the first experiment, the proposed algorithm evaluated the impressions and compared with the algorithm for only TBC. In the second experiment, the proposed algorithm evaluated the effectiveness in practical communications and compared with the algorithm for only DBC.

4.1. An Experiment for Impressions of Robot Responses

To make a comparison with the proposed algorithm, robots A and B are evaluated.

Robot A of the proposed method decides its responses by the combination of dialogues and touch actions. Robot B using only TBC decides its responses by only touch actions as follows:

Case 1) <TOUCHED>

In this case, robot B gives the responses showing its gratitude such as “Thank you for touching me”.

Case 2) <MIS-TOUCHED>

In this case, robot B gives the responses reflecting the <MIS-TOUCHED> attribute such as “That is not the place I want”.

Case 3) <NON-TOUCHED>

In this case, robot B expresses sad intentions such as “Don’t you want to touch me? I’m sad.”

Note that two kind stuffed animals are assigned randomly to robots A and B, respectively, in order to remove appearance impressions of robots.

In the experiment, robots request touch actions and humans’ answers by dialogues and touch actions. To measure impressions of robot responses for all humans’ answers, TC_n ($1 \leq n \leq 13$) is defined. In TC_n , humans indicated their answers. **Table 5** shows the relation of TC_n and humans’ answers.

When humans answer <AFFIRMATIVE> and <NON-TOUCHED>, two situations are assumed as follows: one is humans really have no touch, and another is robots occur touch recognition errors although humans have touch actions.

Therefore, the TC_3 supposes the situation that humans have no touch actions, and the TC_{13} supposes the situation that robots failed to recognize touch actions from humans.

This experiment uses Wizard of Oz experiment [20] to control robot responses for each TC_n . In this experiment, robots are controlled by an experimenter. The experimenter recognizes subjects’ answer attributes (dialogue attributes and touch action attributes), and plays robot voices defined by a response algorithm for each robot.

For all TC_n , the experiment is carried out by the following steps:

1) Answers are provided to a subject.

2) Robot X ($X = A$ or B) requests touch actions.

3) The subject answers.

4) Robot X gives responses.

5) The subject writes answers on a questionnaire sheet about robot X.

6) Repeat steps from 1 to 5 for robot Y ($Y = B$ or A). Eleven students join as subjects of this experiment.

For 6 subjects, X and Y are assigned by A and B, respectively. For 5 subjects, the reversed assignment is carried out.

The questionnaire has 16 adjective pairs in Japanese related to familiarity factors and enjoyment factors with 1 to 7 scales based on the semantic differential (SD) method [21].

Table 6 shows adjective pairs and the related factor of the questionnaire.

Experimental Results

The percentages of total ratings for adjective pairs related to each factor are defined by FFV (Familiarity Factor Value) and EFV (Enjoyment Factor Value). FFV and EFV of robot A and B are calculated for each TC_n .

From SD method [21], the differences of FFV and EFV are clarified by t-test between robots A and B. There is a significant difference if P-values become less than 5%.

Tables 5 and **6** respectively show FFV and EFV for each TC_n , and significant differences of factor values between robots A and B are shown with bold font.

From **Tables 7** and **8**, FFV and EFV of robot A are higher than those of robot B in almost TC_n . In TC_3 , the reason for no difference in FFV and EFV between robots A and B might be that subjects feel something wrong with thankful responses by robot A for touch actions even if subjects have no touch actions.

In TC_4 and TC_{10} , the reason for no difference in FFV and EFV between robots A and B is both robots decide responses by only touch actions.

In TC_{12} , the reason for no difference in FFV and EFV between robots A and B is both robot responses have no relation to subjects’ dialogue answers.

Table 9 shows FFV and EFV for each robot and significant differences for factor values are shown with bold font. In **Table 9**, robot A improves by 20.7 points in FFV, and 12.6 points in EFV compared to robot B. From these results, it is verified that the proposed algorithm is more effective to provide familiarity and enjoyment than the algorithm for only TBC.

4.2. Robot Communication Experiments

In this study, two communications are carried out for 2 minutes by 8 people in ages from 20 s to 50 s perexperiment; the first experiment is for only DBC, and the second is for the combination of TBC and DBC with the robot as shown in **Figure 5**. In this experiment, robots

Table 5. The relation of TCn and humans' answers.

Attributes for touch actions Attributes humans' answers	<TOUCHED>	<MIS-TOUCHED>	<NON-TOUCHED>
<AFFIRMATIVE>	TC ₁	TC ₂	TC ₃ , TC ₁₃
<NOTHING>	TC ₄	TC ₅	TC ₆
<NEGATIVE>	TC ₇	TC ₈	TC ₉
<OTHERS>	TC ₁₀	TC ₁₁	TC ₁₂

Table 6. Adjective pairs and the related factor.

adjective pairs	related factors
(kind, cruel) (Favorable, unfavorable) (safe, dangerous) (warm, cold) (frank, right) (distinct, vague) (accessible, inaccessible) (light, dark) (altruistic, selfish) (exciting, dull) (pleasant, unpleasant) (likable, dislikeable) (interesting, boring) (good, bad) (friendly, unfriendly) (pretty, ugly)	familiarity enjoyment familiarity and enjoyment

Table 7. FFV(Familiarity Factor Value) for each TCn.

	TOUCHED	MIS-TOUCHED	NON-TOUCHED	
	TC ₁	TC ₂	TC ₃	TC ₁₃
AFFIRMATIVE	84.4% (Robot A)	81.3% (Robot A)	71.4% (Robot A)	70.4% (Robot A)
	78.5% (Robot B)	60.2% (Robot B)	66.6% (Robot B)	41.7% (Robot B)
NOTHING	TC ₄	TC ₅	TC ₆	
	87.6% (Robot A) 85.5% (Robot B)	80.4% (Robot A) 55.7% (Robot B)	74.0% (Robot A) 61.5% (Robot B)	
NEGATIVE	TC ₇	TC ₈	TC ₉	
	82.4% (Robot A) 74.5% (Robot B)	79.1% (Robot A) 54.9% (Robot B)	68.7% (Robot A) 60.7% (Robot B)	
OTHERS	TC ₁₀	TC ₁₁	TC ₁₂	
	72.6% (Robot A) 70.7% (Robot B)	71.1% (Robot A) 53.0% (Robot B)	64.6% (Robot A) 62.3% (Robot B)	

Table 8. EFV(Enjoyment Factor Value) for each TCn.

	TOUCHED	MIS-TOUCHED	NON-TOUCHED	
	TC ₁	TC ₂	TC ₃	TC ₁₃
AFFIRMATIVE	80.9% (Robot A)	78.1% (Robot A)	69.6% (Robot A)	67.0% (Robot A)
	73.5% (Robot B)	59.0% (Robot B)	66.0% (Robot B)	39.9% (Robot B)
	TC ₄	TC ₅	TC ₆	
NOTHING	79.4% (Robot A)	73.3% (Robot A)	69.0% (Robot A)	
	79.2% (Robot B)	58.0% (Robot B)	59.2% (Robot B)	
	TC ₇	TC ₈	TC ₉	
NEGATIVE	81.3% (Robot A)	74.6% (Robot A)	67.2% (Robot A)	
	66.4% (Robot B)	57.5% (Robot B)	57.7% (Robot B)	
	TC ₁₀	TC ₁₁	TC ₁₂	
OTHERS	69.4% (Robot A)	67.5% (Robot A)	64.2% (Robot A)	
	67.9% (Robot B)	57.0% (Robot B)	63.8% (Robot B)	

Table 9. FFV and EFV for each robot.

	Robot A	Robot B
FFV	84.7 (%)	64.0 (%)
EFV	78.3 (%)	66.2 (%)

**Figure 5. The scene of TBC with a robot.**

communicate automatically.

elements and interaction components are defined.

The interaction elements are defined by <ROBOT APPROACH>, <HUMAN REACTION> and <ROBOT RESPONSE>. <ROBOT APPROACH> means a question or a touch request from a robot. Figure 6 shows examples of <ROBOT APPROACH> in experiments.

<HUMAN REACTION> means an answer or a touch action from a human. <ROBOT RESPONSE> means a response from a robot. The interaction component is the sequence of interaction elements as follows:

- 1) <ROBOT APPROACH>
- 2) <HUMAN REACTION>
- 3) <ROBOT RESPONSE>

Communications in the experiment are constructed by interaction components. **Table 10** shows the example of a communication.

Table 10. Example of a communication.

1st interaction component	
Hey, touch my arm and let's shake hands.	<ROBOT APPROACH>
(Touch robot's arm)	<HUMAN REACTION>
Thank you! Let's be good friends!	<ROBOT RESPONSE>
2nd interaction component	
Do you like chocolates?	<ROBOT APPROACH>
Yes, I like chocolates.	<HUMAN REACTION>
Chocolates are delicious.	<ROBOT RESPONSE>

For each interaction component, the concentration and expression rates observed from the recorded communications are as follows:

The concentration rate is a standard measurement of a call for attention [22], and it is defined by the following levels:

Level 1: Humans clearly concentrated on interaction component.

Level 0: Humans didn't concentrate on it.

Moreover, human expressions are very useful to measure the interest of interactions with robots [23]. Therefore, expression rate is defined by the following levels:

Level 2: Laugh with opening human's mouth.

Level 1: A smile.

Level 0: Expressionless.

Result and Discussion

First of all, the formal measurements are introduced. Supposed that n is the total number of interaction components. Let i_m ($1 \leq m \leq n$) be the m -th interaction component. Let TI be a set of touch-based and dialogue interaction components. Let DI be a set of dialogue interaction components. X represents TI or DI . $count(X)$ is the num-

ber of interaction components in set X . $CR(i_m)$ and $ER(i_m)$ represent the concentration rate and the expression rate for i_m , respectively.

For i_m and X Increase rate of Concentration rate is denoted by $IC(i_m)$ and Continual rate of Concentration rate, $CC(i_m)$ are defined by the following equations:

$$IC(i_m) = \begin{cases} 1(CR(i_m) - CR(i_{m-1}) = 1) \\ 0(others) \end{cases} \quad (1)$$

$$CC(i_m) = \begin{cases} 1(CR(i_m) - CR(i_{m-1}) = 1) \\ 0(others) \end{cases} \quad (2)$$

In Equation (1), $IC(i_m)$ calculates if current concentration rate $CR(i_m)$ increases from previous rate $CR(i_{m-1})$, and in Equation (2), $CC(i_m)$ checks if current concentration rate $CR(i_m)$ equals to previous rate $CR(i_{m-1})$. Increase rate of Expression rate $IE(i_m)$ and Continual rate of Expression rate $CE(i_m)$ are defined by the following equations:

$$IE(i_m) = \begin{cases} 1(ER(i_m) - ER(i_{m-1}) > 0) \\ 0(others) \end{cases} \quad (3)$$

$$CE(i_m) = \begin{cases} 1(ER(i_m) - ER(i_{m-1}) > 0) \\ 0(others) \end{cases} \quad (4)$$

In Equation (3), $IE(i_m)$ calculates if current expression rate $ER(i_m)$ increases from previous rate $ER(i_{m-1})$ and in Equation (4), $CE(i_m)$ checks if current expression rate $ER(i_m)$ equals to previous rate $ER(i_{m-1})$ and bigger than level 1.

Level Two (2) Expression rate $LTE(i_m)$ is defined by the following Equation which represents the number of expression rate with level 2.

$$LTE(i_m) = \begin{cases} 1(ER(i_m) = 2) \\ 0(others) \end{cases} \quad (5)$$

Supposed that $Y(i_m)$ represents $IC(i_m)$, $CC(i_m)$, $IE(i_m)$, $CE(i_m)$ or $LTE(i_m)$. Common equation $F(X, Y(i_m))$ to measure experimental results is defined by the following equation:

$$F(X, Y(i_m)) = \frac{\sum_{m=2}^{m \leq n} Y(i_m)}{\text{count}(X)} 100, i_m \in X \quad (6)$$

where $\text{count}(X)$ represents a numerator. **Table 11** shows the experimental results of concentration rates where rate[%] represents results of Equation (6).

These results in **Table 11** show that $F(TI, IC(i_m))$ and $F(DI, IC(i_m))$ are low rate and $F(TI, CC(i_m))$ and $F(DI, CC(i_m))$ are more than 90%.

These results are verified that both methods equally have the strong effect to maintain humans' concentration during experiments.

Table 12 shows the experimental results of the expression rates.

From these results in **Table 12**, $F(TI, IE(i_m))$ is 8 points higher than $F(DI, IE(i_m))$. It is verified that the proposed system has higher effect on helping humans to find their interests in communication again than only DBC. $F(TI, CE(i_m))$ is 0.1 point higher than $F(DI, CE(i_m))$. It is verified that the proposed method helps humans to keep their interests in communication. $F(TI, LTE(i_m))$ is twice as high as $F(DI, LTE(i_m))$.

Therefore, it is verified that the proposed method has more effect on giving strong interest to humans than only DBC.

Table 13 shows experimental results of RRT (recognition rate of touch actions) and RRV (recognition rate of voice). In **Table 13**, the total number of touch actions and voices from humans are denoted by TIN and DIN, respectively, and the number of correct recognition of touch actions and voice recognition are denoted by NCT and NCV, respectively.

From results of **Table 13**, it turns out that RRT is 24.5 points higher than RRV.

This result shows that the recognition ability for touch actions is more effective than the voice recognition in the robustness.

5. Conclusions

This paper has proposed an algorithm of combination of DBC and TBC. The proposed algorithm defines rules of system's response considering combination of linguistic answer and touch action from humans.

In order to estimate the proposed algorithm, two ex-

Table 11. Experimental results of concentration rates.

	Σ	$\text{count}(X)$	rate [%]
$F(TI, IC(i_m))$	1	115	0.87
$F(DI, IC(i_m))$	1	129	0.78
$F(TI, CC(i_m))$	113	115	98
$F(DI, CC(i_m))$	127	129	98

Table 12. Experimental results of the expression rates.

	Σ	$\text{count}(X)$	rate [%]
$F(TI, IE(i_m))$	25	115	22
$F(DI, IE(i_m))$	18	129	14
$F(TI, CE(i_m))$	2	115	1.7
$F(DI, CE(i_m))$	2	129	1.6
$F(TI, LTE(i_m))$	12	115	10
$F(DI, LTE(i_m))$	6	129	4.7

Table 13. Results of RR.

Y'	X'	(Y'/X')100
NCT (=28)	TIN (=35)	RRT (=80)
NCV (=161)	DIN (=290)	RRV (=55.5)

periments have been carried out: one is the experiment for impressions of robot responses, and another is the robot communication experiment.

From experimental results, it turns out that the proposed algorithm is more effective to provide impressions of familiarity and enjoyment than the algorithm for only TBC and to help humans to keep their interests in communications than the algorithm for only DBC. Moreover, the recognition ability for touch actions is more robust than for voices. Future works could focus on practical robot development and evaluation for dementia persons.

REFERENCES

- [1] Y. Nakano, J. Hikino and K. Yasuda, "Conversational Agents for Dementia: A Field Study for A Prototype System," *The 25th Annual Conference of the Japanese Society for Artificial Intelligence*, Morioka, 1-3 June 2011, pp. 1-4.
- [2] H. Kozima, C. Nakagawa and Y. Yasuda, "Interactive Robots for Communication-Care: A Case-Study in Autism Therapy," *IEEE International Workshop on Robots and Human Interactive Communication*, Nashville, 13-15 August 2005, pp. 341-346.
[doi:10.1109/ROMAN.2005.1513802](https://doi.org/10.1109/ROMAN.2005.1513802)
- [3] N. Mitsunaga, T. Miyashita, H. Ishiguro, K. Kogure and N. Hagita, "Robovie-IV: A Communication Robot Interacting with People Daily in an Office," *Intelligent Robots and Systems 2006 IEEE/RSJ International Conference*, Beijing, 9-15 October 2006, pp. 5066-5072.
- [4] K. Wada, T. Shibata, T. Asada and T. Musha, "Robot Therapy for Prevention of Dementia at Home—Results of Preliminary Experiment," *Journal of Robotics and Mechatronics*, Vol. 19, No. 6, 2007, pp. 691-697.
- [5] M. Fujita, "On Activating Human Communications with Pet-Type Robot AIBO," *IEEE Proceedings*, Vol. 92, No. 11, 2004, pp. 1804-1813.
- [6] M. P. Bacigalupo, C. Mennecozzi and T. Shibata, "Socially Assistive Robotics in the Treatment of Behavioral and Psychological Symptoms of Dementia," *2006 First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics*, Pisa, 20-22 February 2006, pp. 483-488.
- [7] T. Shibata, "An Overview of Human Interactive Robots for Psychological Enrichment," *Proceedings of the IEEE*, Vol. 92, No. 11, 2002, pp. 1749-1758.
- [8] M. Kanoh, Y. Oida, T. Shimizu, T. Kishi, K. Ihara and M. Shimizu, "Some Facts Revealed from an Elderly and Babyloid Symbiosis," *The 25th Annual Conference of the Japanese Society for Artificial Intelligence*, Morioka, 1-3 June 2011, pp. 1-3.
- [9] N. Futoshi, S. Kazuhiro and K. Kiyoshi, "Human-Robot Tactile Interaction," *Information Proceeding Society of Japan*, Vol. 44, No. 12, 2003, pp. 1227-1232.
- [10] N. Futoshi, S. Kazuhiro, Y. Junji and K. Kiyoshi, "Real-Time Classification of Human Touching Behaviors Using Pressure Sensors and Its Personal Adaptation," *The Institute of Electronics, Information and Communication Engineers*, 2002, pp. 613-621.
- [11] T. Ogawa, K. Morita, H. Kitagawa, M. Fuketa and J. Aoe, "A Study of Dialogue Robot with Haptic Interactions," *7th International Conference on Natural Language Processing and Knowledge Engineering*, Tokushima, 27-29 November 2011, pp. 285-288.
- [12] T. Ogawa, K. Morita, M. Fuketa and J. Aoe, "Empathic Robot Communication by Touch-Based Dialog Interfaces," *Proceedings of 6th International Knowledge Management in Organizations Conference*, Tokyo, 27-28 September 2011.
- [13] R. Freedman, "Using a Reactive Planner as the Basis for a Dialogue Agent," *Proceedings of FLAIRS 2000*, Orlando, 22-24 May 2000, pp. 203-208.
- [14] B. S. Lin, H. M. Wang and L. S. Lee, "A Distributed Agent Architecture for Intelligent Multi-Domain Spoken Dialogue Systems," *IEICE Transactions on Information and Systems*, Vol. E84-D, No. 9, 2001, pp. 1217-1230.
- [15] N. Alm, R. Dye, G. Gowans and J. Cambell, "A Communication Support System for Older People with Dementia," *IEEE Computer*, Vol. 40, No. 5, 2007, pp. 35-41.
[doi:10.1109/MC.2007.153](https://doi.org/10.1109/MC.2007.153)
- [16] A. J. Astell, M. P. Ellis, L. Bernardi, N. Alm, R. Dye, G. Gownas and J. Campebell, "Using a Touch Screen Computer to Support Relationships between People with Dementia and Caregivers," *Interaction with Computers*, Vol. 22, No. 4, 2010, pp. 267-275.
[doi:10.1016/j.intcom.2010.03.003](https://doi.org/10.1016/j.intcom.2010.03.003)
- [17] T. Meguro, R. Higashinaka, K. Dohsaka, Y. Minami and H. Isozaki, "Analysis of Listening-Oriented Dialogue for Building Listening Agents," *Proceeding SIGDIAL 09 Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, Stroudsburg, 2009, pp. 124-127.
- [18] T. Shibata, "Mental Commit Robot for Healing Human Mind," *The Robotics Society of Japan*, Vol. 17, No. 7, 1999, pp. 943-946.
[doi:10.7210/jrsj.17.943](https://doi.org/10.7210/jrsj.17.943)
- [19] K. Wada, T. Shibata, T. Saito and K. Tanie, "Effects of Robot-Assisted Activity for Elderly People and Nurses at a Day Service Center," *Proceedings of the IEEE*, Vol. 92, No. 11, 2004, pp. 1780-1788.
[doi:10.1109/JPROC.2004.835378](https://doi.org/10.1109/JPROC.2004.835378)
- [20] A. Corradini and P. R. Cohen, "On the Relationships among Speech, Gestures, and Object Manipulation in Virtual Environments: Initial Evidence," *In 2002 CLASS Workshop on Natural, Intelligent, and Effective Interaction in Multimodal Dialog Systems*, Denmark, 28-29 June 2002.
- [21] T. Kanda, H. Ishiguro and T. Ishida, "Psychological Analysis on Human-Robot Interaction," *Proceedings of the 2001 IEEE International Conference on Robotics and*

Automation, Vol. 4, 2001, pp. 4166-4173.

- [22] The Press Release of Research Institute National Rehabilitation Center for Persons with Disabilities, "Developed a Robot which Provide Information to Support Independence Action of Persons with Dementia," 2010.
- [23] H. Kuma, "Quantitative Evaluation for 'Laugh' Using a Hobby Robot in Care of Aged Facilities," *The Shinshu Branch of JSLHS*, Vol. 17, 2010, pp. 50-60.
- http://www.rehab.go.jp/hodo/japanese/news_22/news22-03.html